Reel Representation: The Economic Impact of Gender on Bollywood Box Office Revenue

Sidharth Ravi

Professor Genna Miller, Faculty Advisor

Professor Grace Kim, Seminar Advisor

Honors Thesis submitted in partial fulfilment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University

Duke
University
Durham, North Carolina,
April 11, 2025

Acknowledgements

This project could not have been accomplished without the support of so many people. Professor Genna Miller, thank you for your mentorship throughout my Duke experience and helping me find joy while writing this thesis. Thank you for being there for me, hearing me work through so many challenges and for never letting me get in the way of my own success. Similarly, thank you Professor Grace Kim for providing our class a space to work through obstacles and for steering me in the right direction at numerous impasses. Thank you, Professor Michelle Connolly, for helping affirm my methodology and refocus my project overall. I am incredibly grateful for Professor Zinat Aboli of Abingdon and Witney College and Professor Girish Taneja of DAV University, for meeting with me to discuss the specifics of Bollywood Films. All Data has been provided by Nash Information Services LLC, IMDB and Bollywood Hungama.

To my friends, both at Duke and beyond, thank you for graciously offering your time to copy-edit this project (and letting me vent more times than I can count). Particularly, I would like to thank Lauren Nolet, Maya Nair, Ava Wellener, Claire Helgerson, Katie Kotler, Elena Granowsky, Maya Arora, Alexia Jackson and Ella Davis for all your help. Thank you, Sofia Hletko, Morgan Bernstein and Graham Gumbert, for providing critical statistics support throughout this project. You all have made my life, both in college and beyond, more meaningful than I could have ever imagined.

My utmost appreciation goes to my seminar class, especially Brendan Elliott, Karianna Klassen and Kunmi Ojo, for being the best support system throughout this year. From STATA assistance to laughing on all those Zooms, we made it!

In many ways, none of this would have been possible without Deepika Padukone, Sucharita Tyagi and Anupama Chopra for producing films and reviews that inspired a 14-year-old thousands of miles away from Mumbai.

Finally, my family. Thank you, Shreya, for your unbridled honesty. You are the first person I text each morning and the last person I text each night. Thank you for making my life truly joyful. Most importantly, this is for my parents. Nobody in the world is as important to me as you. You both are the people I call in crisis, the first people to read anything I produce, and the only reason I chase my dreams. Words cannot encapsulate my adoration for you both; especially, for seeing the challenge in looking that challenge of raising a kid like me in the eyes and not being afraid of what came. Everything I have and will ever accomplish will be to honor you.

Table of Contents

Abstract

The Hindi Film Industry, known as Bollywood, is seen as a gatekeeper of Indian culture. Annually thousands of films are produced, half a million workers across India are employed and millions in revenue is created. Although Bollywood has ensured increased employment and wage opportunities for women on and off screen, the overall representation of women remains severely low. Little is known about their impact on Bollywood's film revenue. This study uses a novel dataset to estimate the impact of female representation on Bollywood revenue from 2009-2019. We apply a traditional linear regression and use a ratio of female to male characters in a film's cast as a proxy for female representation. Results indicate there is not a significant relationship between an increased female cast composition on box office performance. To check for the diverse impact of star power, I analyzed the gender makeup of the movie star in a film, finding this to have a significant impact on box office revenue. In addition, there is a significant effect of production budgets and genre on box office performance.

JEL Codes: L820, F63, J16, Z11

Keywords: Film Economics, Bollywood, Gender, Female Representation

Introduction

Bollywood, the largest of India's 23 regional dialects by both population and box office revenue (Deloitte, 2016), originated in the late 19th century (Bhaumik, 2005). Today, the total Indian film industry produces between 1,500 and 2,000 films annually (Deloitte, 2016). With national box office revenue reaching nearly $1.5 Billion in 2019 (Shackleton, 2023), representing an overall growth of 11.5% since the 2000s (Deloitte, 2016), the Indian Film Industry employs over half a million workers across India (Maindargi, 2020). The overall value of the film and television industry has nearly doubled since pre-pandemic level (EY, 2024) while Bollywood-specific box office revenue has fluctuated since the introduction of online streaming platforms (Jha, 2023).

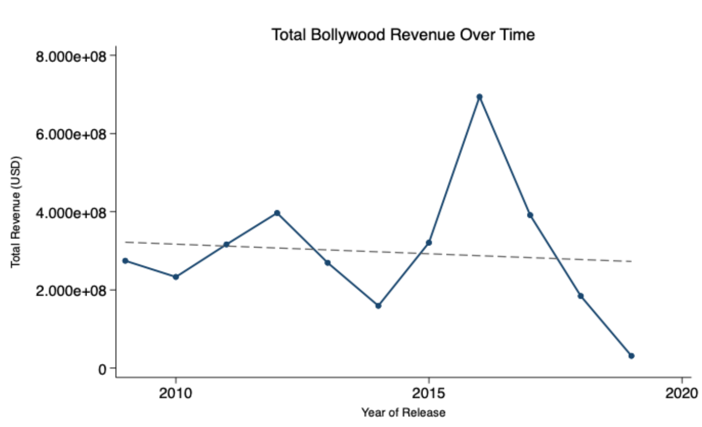


Figure 1: Yearly Bollywood Revenue since 2019

One prominent critique of Bollywood is its stereotypical portrayal of female characters and the limited representation of women, with most storylines centering male protagonists prior to the turn of the 21st century (Khadilkar, et al., 2022). For example, female actresses are frequently depicted in dance sequences designed to please male leads or act as romantic love interests, yet the central narratives focus on the "hero's arc". Cinema has long been on the cutting edge of social progress (Davids, 2023) and with evolving societal expectations, more women are in front of and behind the camera (Khadilkar, et al., 2022). Many prominent actresses have made strides to improve this (Zaidi, 2021), facing difficulty in procuring budgets and industry respect, leading representation of women in cinema to stagnate at less than 30% (Lingam et. al, 2023).

This study aims to analyze the representation of women in Bollywood through an economic lens. The research questions are as follows: **What economic impact does female representation have on Bollywood box office revenue?** and **Does the impact of female representation differ between Indian and North American markets?** This study quantifies female representation on screens by both analyzing the gender of the leading movie star in a film and the ratio between named female and male characters.

The field of "film economics" has traditionally focused on determinants of box office success in Hollywood with several studies analyzing the filmmaking process. However, applying this work to Bollywood has only been done a handful of times due to limited access to verifiable data. By using a novel dataset, this study introduces gender as the primary independent variable with several other explanatory variables to adhere to prior literature. Notably, gender is viewed as a binary between female and male for replicability. Further, viewing Indian and American box office characteristics separately illustrates vital differences in global film consumption. Finally, the response variable in this study is box office revenue as opposed to other gauges of growth to emphasize the demand for an individual movie rather than the overall industry. This approach has been modeled in numerous studies focused on Hollywood (Smith et al, 2010) and has been implemented in the Bollywood context (Fetscherin, 2010).

This paper begins with an overall review of existing literature, both in Hollywood and Bollywood, regarding determinants of film success including gender. Then, I highlight the theoretical framework and dataset I used to complete analysis. Next, I affirm the empirical methodology before diving into results. Finally, I highlight the limitations of this study and point to areas for future research.

Literature Review

Many have conducted studies to better understand predictors of box office revenue. While gender is a more recent variable of interest, only appearing in studies starting the mid-to late 2010s, other variables are still relevant in order to better understand the landscape of film economics. To that end, literature broadly falls into two classifications: macroeconomic and microeconomic (McKenzie, 2012). The purpose my study is to better understand the impact of diverse gender representation on each individual film's box office revenue. Thus, microeconomic literature was influential for my research. The following two subsections detail relevant literature. The first, emphasizes all papers pertaining to determinants of box office revenue noting those which utilize Bollywood data. The second, examines papers which suggest that gender influences box office revenue and the tools scholars have used to quantify this relationship.

Film Economics

While exploring determinants of film viewership, researchers typically analyze box office revenue. Some alternatives include surveying online platforms, such as the Hollywood Stock Exchange, to gauge real-time online user interaction supporting certain films (Elberse et. al, 2003). Others use internet search results, especially as more films appear on digital platforms such as Netflix or Amazon Prime (Feng, 2017). By in large, the most common estimate for the demand of an individual film is by its overall box office revenue.

Many studies focus on predicting box office returns by highlighting various factors influencing film viewership. For example, Prag & Casavant found evidence indicating a significant relationship between genre, star presence and age certifications (Prag et al, 1994). This study relies on a dataset of 600 films in 1990. Therin, researchers compiled box office receipts and other production costs found in Variety Magazine, a prominent trade publication. This study finds that production costs, Academy Awards, movie stars and genre significantly positively influence box office revenue. Others highlight a positive relationship between critic's reviews (Eliashberg et al, 1994), awards and other nominations (Deuchert et al, 2005), and sequels (Bohenkamp et al, 2015).

These studies rely on Hollywood data, typically accessed from the Internet Movie Database (IMDb) or other trade publications. This remained consistent even while looking at international markets and non- United States determinants of film success such as colorism. Colorism, in this case, referring to racial preferences for fairer-skinned individuals in China (Hermosilla et al, 2018). IMDb was used to provide casting and genre information within the dataset. After procuring basic information from IMDB, researchers augmented with other individualized characteristics of films. Hermosilla and colleagues found that colorism influences films especially those marketed towards Chinese audiences. While only researching Hollywood projects, this study provides useful context for prior data sources common in the field.

Genre has remained a consistently studied film revenue predictor. For example, a study published in 2009 used top grossing films released between 1997 and 2001. In all, the dataset comprised over 400 film observations and included year fixed effects. Additionally, this study specifically emphasized genres with wide ranging appeal. As a result, the variables of Comedy and Horror, positively influenced box office returns (Brewer et al, 2009). Much like other studies, Genre was considered a critical determinant of box office revenue.

The impact of movie stars, on the other hand, has been volatile: indicating changes in industry practice. This is due to their overall importance on the filmmaking process (Treme et al, 2018). Stars are used as a means of risk reduction for filmmakers (McMahon, 2023). McMahon collected IMDb data and found that stars typically produce larger release schedules leading only a handful of successful actors to overwhelm box offices. This creates larger inequity within the industry. McMahon's dataset comprises Hollywood films released between 1983 to 2009, utilizing IMDB casting information as a proxy for relative importance on film narrative. This study found that after 1980, there have been more independent and smaller distributors leading to fewer actors seeing theatrical releases. Other studies utilize year end lists to approximate overall notoriety (David et al, 2023).

When analyzing critics reviews researchers used publicly available data through IMDb user generated ratings. Berg, for example, focuses on the day a film was released, budget, and film perception by both critics and audiences. Researchers compiled box office revenue data on over 5,000 films, combining this with available critics and user feedback on IMDB and Metacritic. Subsequently, researchers employed both a month and year fixed effect to capture seasonality and yearly differences. This study highlighted that critics reviews were relevant to box office success only if the film had relatively smaller budgets within the dataset. Additionally, audience perceptions differ significantly from critical acclaim (Berg & Raddick, 2016). Basuroy, alternatively, compiled a random selection of 200 films cross referencing these with results from trade publications and newspapers. He finds that both positive and negative reviews act as influencers of box office revenue (Basuroy et al, 2003). This provides useful information for sourcing critics reviews.

To some extent, most studies focused on microeconomic influences of box office revenue include variables for genre, star presence, quality indicators/reviews, budget and seasonality (McKenzie, 2012).

Data management has been a focus for researchers in film economics often dealing with heavy-tailed box office information. The overall distribution of box office revenue is highly skewed with a few films making much of total revenue (Collins, et. al, 2002). As a result, many studies feared heteroskedasticity and unbounded variances with results that are applicable only to specific datasets. Little consensus has been reached (McKenzie, 2012) with almost a dozen papers highlighting different mechanisms of approximation. Some papers employ what is known as a "quantile" or "threshold" approach (Hoffmann et al., 2017). In other words, by analyzing films based on their box office sizes each group is distinct. However, most papers employ a logarithmic transformation of the respective dependent variables to then analyze box office revenue (Feng & Sharma, 2016). De Vany and Walls, for example, utilized log-linear regressions and probability distributions (De Vany & Walls, 1999). This study, one of the first in the field, analyzes nearly 2000 films. They found that converting box office revenue to a logarithm increases the predictive power of the Ordinary Least Square regression. My dataset undergoes the same transformation.

Fetscherin was the first to explore determinants of box office revenue in Bollywood. Therein, he utilized sales data from the United Kingdom and found little association between genre and star power on the box office (Fetscherin, 2010). Fetscherin uses British box office receipts to gauge film success, particularly in the United Kingdom, creating a dataset of 300 films. Within it, he separates film specific attributes finding reviews and director popularity to increase box office revenue. Aboli, on the other hands, uses data scraping of IMDB to compile film revenue for all films released in the 2000s. She then uses online publications such as Bollywood Hungama to estimate production budgets, critics reviews, and star presence. She finds those same variables – production budgets, critics review and star presence - to be most predictive for the Indian box office (Aboli, 2022). Dastidar, on the other hand, focuses on Chinese film markets. His dataset comprises 245 Bollywood films released in China. After employing an instrumental variable approach, he finds critics reviews to not significantly impact box office revenue (Dastidar, et al., 2020). Niraj and Singh highlight this relationship in their study tracking 48 films released in 2011. They then analyze audience discussion of these films on social media platforms. After comparing these results to the Times of India, the largest newspaper in India, they found significant relationships between critic's review and online reviews on box office revenue (Niraj &

Singh, 2015). The difference in these results highlights the need for replicated econometric analysis.

Economic analysis has been sparse in Bollywood due to limited data availability (Dastidar, et al., 2020). This leads many scholars to rely upon various case studies of qualitative factors (Patil, et al., 2022) as opposed to larger empirical studies. Most, if not all, rely on publicly available data through IMDb and other online trade publications such as Bollywood Hungama. In our study, we rely on a novel dataset from Nash Information Services LLC while using other sources for indicator variables.

It is crucial to emphasize that gender has been overlooked from an empirical economic lens in the literature thus far beyond the mentioning of a film's central protagonist (Aboli, 2022).

Gender Representation in Film

Applying gender as an independent variable of interest with respect to film revenue has been largely understudied across Film Economics. Partly in response to evolving industry standards (Treme et. al, 2018) more scholars have begun analyzing gender. As a result, there is no standard practice for quantifying representation. The literature illustrates three categories for analyzing gender in films: the presence of female movie stars, narrative representations, and Female-To-Male ratios. My study does not highlight other important crew or production staff, such as the gender of a film's director or writers, though this is potential area for future research.

Presence of Female Stars

The impact of movie stars on box office returns has been studied since Paul and Casavant's seminal study in the 1970s. Due to their importance on every stage of the filmmaking process, a movie star's influence on box office returns is undeniable. This analysis can be applied to gender, as the gender of a film's central character can act as a gendered differentiator. For example, Treme compiles 275 films released between 2004 and 2006. She then analyzes male and female stars, alongside budget, rating and genre. She then finds female stars do not positively impact box office revenue (Treme et. al, 2018). Notably, she does not account for more than one star appearing in a film's cast. However, Aboli finds that female stars positively impact box office revenue in Bollywood (Aboli, 2022). Like Treme, she does not account for the presence of both male and female stars in a film. This difference illuminates distinct audience relationships with movie stars. Bollywood, and the overwhelming presence of Romantic Comedies, is an interesting case study on the difference gender plays for actors (Aboli, 2022). I have incorporated the gender of a movie star into my analysis.

Narrative Representation

This approach is more complex and questions the quality of representation in films (Nguyen, 2023). For example, "Million Dollar Maybe? The Effect of Female Presence in Movies on Box Office Return?" (Lindner et al., 2015) uses the Bechdel test as a proxy for female representation in movies. The Bechdel Test is a three-question test used to examine female representation in a film and has been used for decades. Lindner accessed United States domestic box office performance from IMDB and curated the films that were released in theaters. In final regressions, she controlled for production budget, genre, star quality, and aggregate critical reception. In all, she finds that films that successfully completed the Bechdel Test negatively influenced box office revenue. Concurrently, she finds that many films that pass the Bechdel Test are associated with lower production costs.

Other papers highlight the Bechdel Test's impact on international box office performance of Hollywood films (Valentowitsch, 2023). While using panel data of 515 randomly selected films across many countries he finds a significant positive relationship between successfully passing the Bechdel Test and international box office revenue. Further, this relationship is dependent on the socioeconomic development of selected international markets. The data source of this study was Box Office Mojo and IMDb, controlling for genre, production budget, and runtime. This paper provides insight of a mechanism to factor gender to ultimately analyze box office revenue. The Bechdel Test is appropriate for Hollywood and there are many online resources compiling information. Hindi cinema historically limits portrayal of women resulting in only a handful of films that complete all elements of its criteria (The Ideas Lab, 2018). Thus, it was not used in my study.

Female to Male Ratios

The third approach requires analyzing the composition of an individual film, either narratively or cast makeup, to estimate the representation within the film. To understand the overall quantity of female representation in movies, some work has been done to analyze the dialogue between characters. In a study by Nguyen, researchers compared The Bechdel Test and the percentage of female spoken dialogue finding higher efficacy in quantifying representation (Nguyen, 2023). Nguyen's study uses machine learning and analyzes over two thousand films released between 1960 and 2018. He finds that by simply having one conversation between two women during a film's runtime, film revenue increased by 27%. Additionally, the type of representation was equally as important for audiences. Others focus on the ratio between male and female characters in movies. This work has been repeated across disciplines. One study focuses on G-Rated movies released in the United States between 1990 and 2005 (Smith et al, 2010) finding that male characters outnumbered female characters by a ratio of 2.57 to 1.00. This study utilized the top 100 G-rated films released from 1990 to 2005, finding that the personality distinctions between characters impacted viewership. Another study in 2020, utilized female cast ratio to approximate gender representation. This dataset was created by using machine learning to scrape IMDB and TheBechdelTest.Com, an online database, and had analyzed nearly 8,500 total observations. Researchers compared female-to-male ratio to budget size alongside box office revenue finding many films lack budgets as representation increases causing lower box office returns. They further found the female cast ratio to more accurately gauge female representation and a stronger predictor of box office success (Yang et al., 2020). My thesis applies this method to Bollywood films, finding similar results.

Theoretical Framework

This section offers an overview of the foundational theory of my study, the basis for the female to male cast ratio, and the importance of chosen control variables used in final empirical specifications.

Determinants of Box Office Revenue

This study exploits the concept of "hedonic demand" (McKenzie, 2022) which decomposes a "good" (in this case an individual film) into factors to better predict audience preferences. It emphasizes that people watch films because of specific characteristics. This study argues that the box office revenue, demand, for a film is in response to various factors including gender representation.

Studies began viewing film revenues as products with different characteristics beginning the late 1980s (Ravid, 1990). I have outlined each control variable and corresponding theory in this section. As previously stated in the literature review, this approach has been used for many decades. Additionally, this method has evolved past tangible properties of a film and explores factors such as race (Hermosilla et. al., 2018) and shared lived experiences/social learnings (Moretti, 2008). Today, scholars argue that as films align with audience lived experiences films have stronger box office outcomes. I will be viewing gender representation as an individual film characteristic and exploring impact on box office revenue.

Female-to-Male Ratio

A key variable in this study is the ratio between named male and female cast members in a film. This proximate gauge for gender representation utilizes properties of two well-studied tests: The Bechdel Test and the Female Face Ratio (FFR). The former, created by Allison Bechdel, is a simple mechanism for qualifying gender diversity within films (Bechdel, 1986). The prongs of the test are as follows:

1. Does the film possess more than one named female character?
2. Do the two, named, female characters engage in conversation?
3. Is there conversation between two named female characters about something other than a man?

There are numerous shortcomings with applying this theory to films. Firstly, it does not fully encapsulate the breadth of diverse representation. This theory argues that any conversation between two women makes a film inherently diverse. Secondly, there has been an increasing trend in creating artificial moments in film simply to "pass" The Bechdel Test (Treme et al, 2018). However, numerous theaters and critics utilize The Bechdel Test as a barometer for gender representation (Lindner et al, 2015). The first prong of this test has been combined with the "Female Face Ratio" (FFR). As the name suggests, the FFR analyzes the ratio of female to male faces in each film (Mazieres et. al, 2021). This metric has been used to understand the impact gender has on children's films (Smith et al, 2010). As previously stated, compared against the Bechdel Test this ratio (Yang et al, 2020) better represented gender in film. As the primary goal of the original Bechdel Test is simply whether two named female characters engage in conversation it hides the impact of the aggregate gender of the cast. By using an element of The Bechdel Test and identifying a ratio, we can understand the impact of a film's overall cast on box office revenue.

Control Variables

Many papers and studies surrounding films, especially those examining aggregate demand and consumption of film revenue, control for variables which otherwise impact box office data. Briefly, this subsection outlines those variables and their expected relationship with box office revenue.

Critics and Reviews

Many, if not all, exploratory studies modeling the demand for films pay special attention to reviews (McKenzie, 2022). This work includes the impact of both critics and popular reviews to highlight potential differences in quality indicators. While these studies have infrequently focused on Asian film industries, they provide valuable insight into relevant theories and methodologies. For example, a 2005 study using a difference-in-difference test found that positive reviews have a substantially large impact on the box office revenue of films (Reinstein & Snyder, 2005). Other studies have used awards nominations (Ginsburgh, 1999). In my study, I utilized *Times Of India's* aggregate critic's score as a proxy for the overall critic's reception. As the Times of India is the largest newspaper in India audience members are likely to respond to these reviews. To capture audience perspective on a particular film, a variable has also been included to capture online reviews from IMDB. As mentioned in the literature review, these two variables are distinct and articulate dissonance between critics and audience perception.

Production Budget

Production budgets are a crucial predictor of film success and have been studied by various film economists, especially as they influence much of the filmmaking process. For example, a recent study found that of several determinants of box office revenue (Hao, 2023) production budgets significantly impacted revenue. These were based off scraping online data sources using machine learning ultimately compiling over 10,000 observations. Others establish co-financing as a proxy for the size of various projects, also showing strong impacts on box office revenue (Palia, 2008). Similarly to box office revenue, production budgets also follow a heavy-tailed pattern (Brewer et al, 2009). To compensate, many papers apply the natural log of available budget details to better fit the available dataset. In this study I rely on publicly available information from Bollywood Hungama, an important trade publication, to estimate the production budget sizes. Aboli replicates this approach in her work as they approximate audience interpretations for a films scale (Aboli, 2022).

Seasonality

Seasonality captures the impact certain times of the year have on a film's success. For example, if a film were to be released near Eid or Holi, that film would likely see higher box office returns than a film released in the middle of the year (Aboli, 2022). Most, if not all studies on the determinants of box office revenue control for a variable regarding the time of release. This variable controls for pre and post release time effects which may influence a movie's marketing and distribution. To do so, some studies create between four-to-five-week intervals to establish seasonality (Einav, 2007). This approach is like Aboli's study wherein researchers established a variable indicating the month of a film to control for underlying seasonality (Aboli, 2022). My study does the same by controlling for the month of a film's release.

Genre

Every empirical study analyzing the characteristics of box office revenue controls for Genre due to its insurmountable impact on viewership. The justification is relatively simple as individuals find specific types of films appealing (Wühr et al, 2017). As stated in the literature review, Prag & Casavant found that while inconsistent, genre was an important factor in determining a film's success (Prag & Casavant, 1994). Compared to the Bollywood context, Genre had less significant an impact on box office revenue. Some find Drama are impactful (Aboli, 2022) while others highlight the importance of comedy films (Dastidar, et al., 2020). I control for Genre using classifications provided by Opus Data and condense to 5 broader categories, as to adhere to prior literature.

Stars

No consensus has been reached to operationalize "star power" due to its importance on box office revenue. That said, most studies consider its overall impact when studying box office success (McKenzie, 2022). Some studies utilize award wins to approximate an actor's relative "star power" (Basuroy et al, 2003) others utilize annual lists by trade publications (De Vany & Walls, 1999). Both studies find that star presence positively impact overall box office revenue, though the size of this impact varies based on individual data. The present studies utilize a list by the trade publication Bollywood Hungama, highlighting the top 25 actors/actresses, to properly account for extraneous factors such as brand sponsorships influencing audience perception of an actor or actress. Further, to attribute gender as others have done (Treme et al, 2018), this study controls for the gender of a star to better specify audience perception.

Sequels

Sequels and/or Franchises have been studied for decades. Theoretically, priming an audience for a film's characters, themes, or narratives provide success and lower risk for filmmakers (Palia, 2008). Sequels act as brand extensions with a film's original brand carrying weight for further consumption (Kim and Kim, 2018). Most studies control for this and in the Bollywood context there was a significantly positive relationship between sequels and box office revenue (Dastidar et al, 2020). My study creates a dummy variable, indicating whether a film was or was not a sequel or part of a franchise.

Data

Several studies focusing on Bollywood highlight data validity as a consistent issue (Aboli, 2022). Data is both inaccessible and unverifiable due to a lack of centralization within the film apparatus. This leads to a standstill in terms of analytical capabilities. This section outlines the sources I used to complete my analysis, the appropriate dependent, independent and control variables, summary statistics and correlation matrices. I finish this section by illustrating the potential limitations in my data and how this influenced the results I derived.

Sources

This study utilizes data obtained by Opus Data; a database owned by Nash Information Services LLC. This company provides box office revenue, genre information, and relevant movie characteristics to numerous online platforms such as The Numbers. Having been used by various other economics studies, Opus Data claims accuracy within 1 - 10% of the stated value. In that accuracy comes a lower number of films. Prior to 2000, OpusData has reported losing validity (Lim, 2016). As such, the selected period for this study is between 2009 and 2019. This pre-pandemic period was during established times of development within Bollywood (Jha, 2019). From the genre classifications provided by Opus Data, "5" prominent genres were used in final regressions: Action, Comedy, Drama, Romantic Comedy and Thriller/Suspense. 13 Films fell into other categories. IMDB was used to verify the film genre of these projects.

Other Data Sources, including Box Office Mojo, Box Office India and IMDb were used to complete information regarding production budget. As replicated by Aboli, this is the closest estimation to production budget over time. While these sources do not report accuracy they provide an estimation in terms of audience perception. Especially as this study focuses on the demand for film, what an audience assumes of a production budget provides insight in terms of how it is perceived.

This dataset, capturing estimated revenue, production budget, genre, sequel and distribution house was then augmented with information from IMDb, Times of India and Bollywood Hungama. Production budget was converted from INR to USD using the Wall Street Journal January 2018 conversion rate. IMDb has been used across film economics to identify audience perception of a film based upon its audience film rating. Additionally, IMDb was used to calculate the ratio of named female to male cast members. To do this, I manually counted the number of named female and male characters present within a film replicating prior work. I then verified the gender of given cast members based upon IMDb listed character names and story synopsis. To verify actors or actresses not easily identifiable, I accessed appropriate Wikipedia pages to discover gender presentation. In all, I repeated this process for the 235 films in the final dataset. Subsequently, I created a new variable with a simple ratio of female to male cast members. To capture critic's review, I used the Times of India. As the nation's largest newspaper this was an appropriate quality indicator. Like the process for ratio, I manually entered the corresponding critics score in the form of a decimal for each film. Finally, Bollywood Hungama, a trade publication, was used to identify star actors and their overall relevance on the Indian market. I used IMDb casting information and inputted any actor which appeared in the top 25 of this list, marking whether there was a male star, female star, or both. After doing this, I created four dummy variables to capture the variety in star presence indicators within the dataset. In all, this process resulted in a complete dataset of 235 films.

Variables and Data Information

The final dataset included 235 films released in India. While significantly less than Aboli's final dataset, I prioritized data accuracy to complete my analysis. The following tables outline the various variables I used in my study, their sources, and their operational definition.

Table 1. Dependent Variable

Variable	Definition	Source
log(revenue)	A given film's domestic box office revenue. All values are in log 2018 US dollars.	IMDb, Opus Data

Table 2. Independent Variable

Variable	Definition	Source
Ratio	The ratio between named female to male cast members in a film.	IMDb

Table 3. Control Variables

Variable	Definition	Source
log(Production Budget)	Production budget of a film excluding markeitng and advertising costs.	Box Office India, Box Office India
Star Presence	Bollywood Hungama ranked an actor's overall star quality based upon endorsments, past box office success, and social media presence. Four Dummy Variables were created: No Star, Only Male Star(s), Only Female Star(s), Both Male and Female Stars.	Bollywood Hungama
Release Year	Dummy Variables were created to account for any changes year over year.	Opus Data
Genre	Dummy Variables were created to control for each genre. The genre list is as follows: Action, Comedy, Drama, Romantic Comedy, Thriller/Suspense	OpusData
MovieMonth	Month of release was utilized to capture seasonality.	OpusData
Popular Review	The audience perception of a film on a 10 point scale, treated as a continuous variable in the final regression.	IMDb
Critics Review	Critic's perception of a film on a 5 point scale, treated a continuous variable.	Times of India
Sequel	Whether a film was or was not a sequel. 1 = Yes 0 = No	OpusData

Table 4. Summary Statistics

Variable	Mean	Std. Dev	Min	Max	Observations
log(Indian Revenue)	14.864	2.261	6.217	19.495	235
log(US Revenue)	13.305	1.448	8.542	16.759	235
Ratio	0.577	0.490	0	4	235
log(Production Budget)	15.527	0.800	12.442	18.1	235
Star Presence					
No Stars	0.319	0.467	0	1	235
Has Male Star	0.255	0.437	0	1	235
HasFemale Star	0.157	0.365	0	1	235
Has Both Male/Female Star	0.268	0.444	0	1	235
Release Year					
2009	0.115	0.320	0	1	235
2010	0.094	0.292	0	1	235
2011	0.081	0.273	0	1	235
2012	0.089	0.286	0	1	235
2013	0.068	0.252	0	1	235
2014	0.077	0.267	0	1	235
2015	0.149	0.357	0	1	235
2016	0.098	0.298	0	1	235
2017	0.077	0.267	0	1	235
2018	0.089	0.286	0	1	235
2019	0.064	0.245	0	1	235
Genre					
Action	0.226	0.419	0	1	235
Comedy	0.226	0.419	0	1	235
Drama	0.285	0.452	0	1	235
Romantic Comedy	0.175	0.380	0	1	235
Thriller	0.089	0.286	0	1	235
MovieMonth					
Jan	0.077	0.267	0	1	235
Feb	0.094	0.292	0	1	235
March	0.068	0.252	0	1	235
April	0.102	0.303	0	1	235
May	0.047	0.212	0	1	235
June	0.094	0.292	0	1	235
July	0.090	0.286	0	1	235
August	0.102	0.303	0	1	235
September	0.123	0.330	0	1	235
October	0.098	0.298	0	1	235
November	0.081	0.273	0	1	235
December	0.072	0.260	0	1	235
Popular Review	6.026	1.423	1.8	8.4	235
Critics Review	0.665	0.112	0.4	0.9	235
Sequel	0.115	0.320	0	1	235

From a descriptive standpoint, the average ratio between female to male named cast members is 0.577. This translates to mean that on average for every one female named cast member, there are roughly two named male cast members. The distribution of films released are consistent within this period with Dramas comprising roughly 29% of films released. Further, only about 16% of films had a solo female star compared to 26% which had a solo male star. However, most films had both female and male stars or no star. The distribution of release years is similarly consistent, with the highest from 2015 and the lowest from 2019. By controlling for release year, any extraneous industry wide or larger economic conditions will not impact the regression estimations. The average popular and critics review align at around 0.6 indicating that audiences and critics enjoy most films released more than a neutral position. Finally, only 12% of films released in this period were sequels.

Figure 2. Indian Box Office Revenue

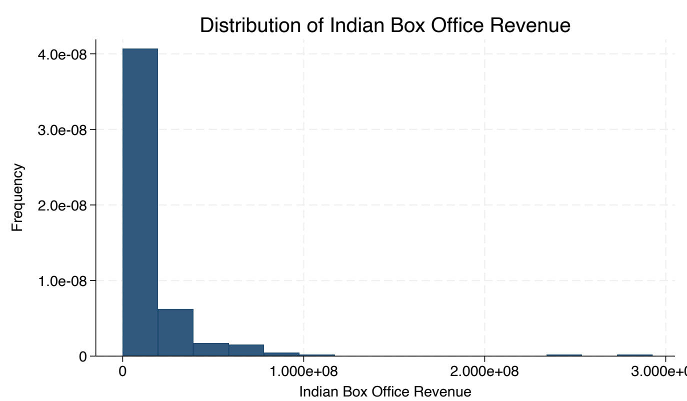
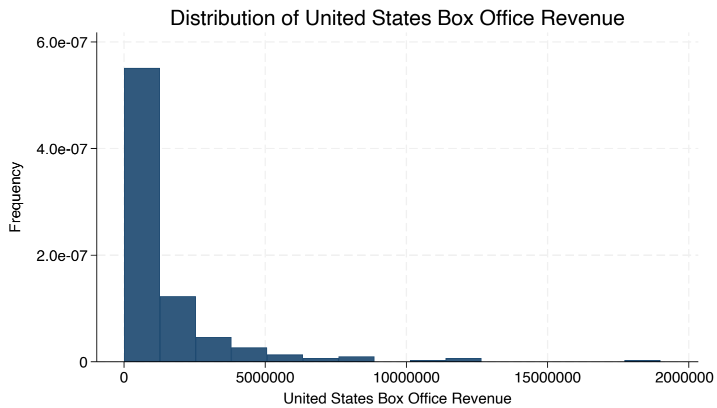


Figure 3. United States Box Office Revenue



As seen in Figure 2 and 3, the distribution of box office revenue represents a heavy-tailed right skewed plot. This aligns with prior literature (McKenzie, 2012). To properly handle this dataset, a log was taken of the dataset to ensure normality within the data plots. Those distributions are below.

Figure 4. Log Transformed Indian Box Office Revenue

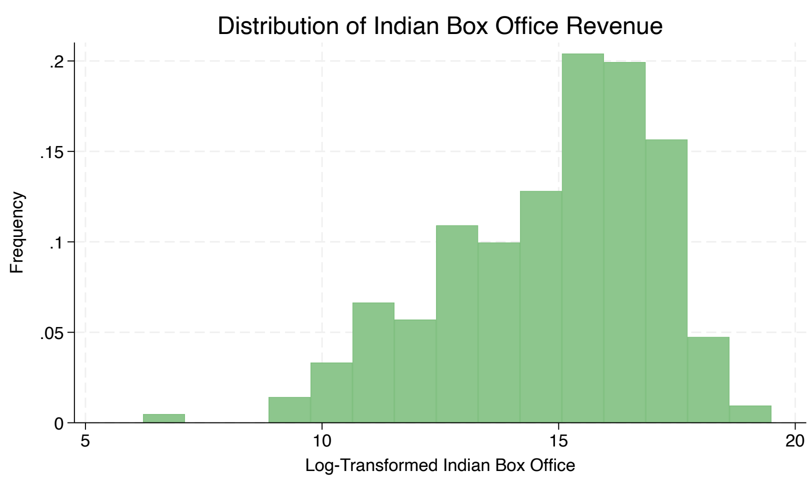
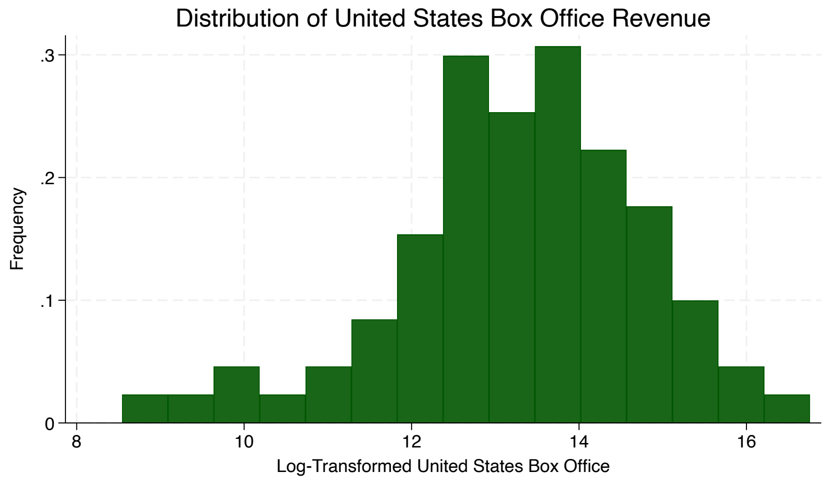


Figure 5. Log Transformed United States Box Office Revenue



Compared to original distribution; by transforming the data the overall distribution better adheres to normality. Although Indian Box Office Revenue still has some skewness, likely attributed to the size of the overall dataset, to adhere to prior work done the log-transformed versions of revenue were used for remaining analysis. Table 5 is a correlation matrix with all independent and control variables.

Table 5. Correlation Matrix

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	log(Indian Revenue)	1.0000													
2	log(US Revenue)	0.5229	1.0000												
3	Gender Ratio	0.0043	0.0336	1.0000											
4	log(Production Budget)	0.2495	0.6075	-0.1248	1.0000										
5	No Star	-0.3373	-0.5024	-0.0714	-0.4580	1.0000									
6	Has Male Star	0.0867	0.0668	-0.0859	0.1890	-0.4060	1.0000								
7	Has Female Star	-0.0714	0.0054	0.1117	-0.1421	-0.2967	-0.2519	1.0000							
8	Has Both Male and Female Star	0.3305	0.4621	0.0690	0.4118	-0.4149	-0.3522	-0.2574	1.0000						
9	Release Year	-0.3025	0.2027	-0.0409	0.3916	0.0216	-0.0268	0.0833	-0.0648	1.0000					
10	Genre	-0.1294	-0.0669	0.1389	-0.2710	0.0502	-0.1312	0.1504	-0.0468	-0.0205	1.0000				
11	Movie Month	-0.0724	-0.1142	0.1682	-0.1755	0.0756	-0.1106	0.1546	-0.0976	-0.0259	0.1999	1.0000			
12	Popular Review	0.0586	0.1867	0.0764	-0.2275	0.0900	-0.0504	-0.0131	-0.0348	-0.0591	0.1334	0.0265	1.0000		
13	Critics Review	0.0977	0.2942	-0.0243	-0.0240	-0.0332	-0.0219	0.0024	0.0548	-0.0205	0.1340	-0.0463	0.6138	1.0000	
14	Sequel	0.0861	0.1733	-0.0994	0.1956	-0.0208	-0.0279	-0.1169	0.1457	0.0095	-0.1683	-0.0796	-0.2002	-0.0105	1.0000

Table 5 shows that the highest correlations are intuitively associated. These include log(Indian Revenue) and log(US Revenue) with p = 0.5229, log(US Revenue) and log(Production Budget) with p = 0.6075, and various star power indicators with budget p = 0.4118. Beyond this, there is a strong correlation between popular review and critics review p = 0.6138. This is typically seen in Film Economics (Aboli, 2022). Further, checking variance inflation factors from regression equations showed no sign of multicollinearity. Beyond these notables, no correlations exceeded 0.25.

Data Limitations

The largest limitation for this dataset is its size leading to a narrower lens of the Bollywood film market. This is a trade-off between accuracy and quantity. Rather than fully augment the dataset with information from IMDB or Bollywood Hungama, this dataset prioritizes authenticity of Box Office data.

Secondly, this dataset lacks information on marketing spend for a given film. While the production budget has been seen as a proxy for this (McKenzie et al, 2022) it is not without shortcomings. External media presence, such as paparazzi, is not accounted for in this dataset.

Thirdly, this dataset cannot fully capture the effect of streaming services on box office revenue. By employing a year fixed effect, some of this volatile has been mitigated. As a result, the number of users who had access to the internet and other streaming services increased. However, in final empirical specifications, it was not appropriate to control for this specific year as fluctuations appear yearly starting 2015. For consistency, a year fixed effects as implemented to account for all economic and extraneous factors. However, there still might be unknown economic influences not captured by this dataset.

Empirical Specification

This section outlines the specific methodology, Ordinary Least Squares (OLS) regressions and results of this study. The primary equations are the same regardless of whether applied to an Indian or United States context however both are listed separately below. This has been done in other studies looking at Bollywood in foreign markets, such as the United Kingdom (Fetscherin, 2010).

Methodology

Indian Market

The initial regression used in this study followed a thorough review of the literature and crucial variables to include. As previously mentioned, most studies create log-linear regression. The estimated regression equation, prior to removing insignificant findings, was the following:

$$\begin{aligned}\log(IndianRevenue_{it}) = \; & \beta_0 + \beta_1 \cdot Ratio_{it} + \beta_2 \cdot \log(ProductionBudget_{it}) + \\ & \beta_3 \cdot StarPresence_{it} + \beta_4 \cdot ReleaseYear_t + \beta_5 \cdot Genre_{it} + \beta_6 \cdot MovieMonth_{it} + \\ & \beta_7 \cdot PopularReview_{it} + \beta_8 \cdot CriticsReview_{it} + \beta_9 \cdot Sequel_{it} + \varepsilon_{it} \end{aligned} \quad (1)$$

In this equation, the subscript "i" indicates individual films and within "t" years. $Release\ Year_t$ refers to year fixed effects for this equation between 2008-2019. β_0 refers the baseline logged Indian box office revenue without any changes in determinants of film revenue. β_1 is the effect changes in the ratio of female-to-male cast members have on the elasticity of Indian Revenue in terms of percent. β_2 refers to the elastic change between production budget and Indian revenue in percent. The Variable "Star Presence" is a dummy variable with four sub-categories, presence of a male star, presence of a female star, presence of both a male and a female star, and the presence of no star. β_3 is the percent change in Indian Revenue due to the presence of stars. β_5 refers to any elastic changes to Indian film revenue that are related to the Genre of a film, for which 5 dummy variables were created: Action, Comedy, Drama, Romantic Comedy and Thriller. Likewise, β_6 is the percent change on Indian box office revenue based upon the month a film was released. β_7 and β_8 respectively are the impact of Popular and Critics Review, both of which are the fractions against the maximum score. These coefficients are the impact of one-unit-increase or decrease of the popular and critics review of a film. β_9 is the elastic impact the film being a sequel. Finally, ε_{it} refers to the uncorrelated error term of the regression.

North American Context

Adhering to prior literature (Fetscherin, 2010), the fundamental regression equation is the same between the Indian and North American context. The interest here is how the explanatory variables differ between two markets. As such the estimated regression equation is as follows:

$$\log(USRevenue_{it}) = \beta_0 + \beta_1 \cdot Ratio_{it} + \beta_2 \cdot \log(ProductionBudget_{it}) + \beta_3 \cdot StarPresence_{it} + \beta_4 \cdot ReleaseYear_t + \beta_5 \cdot Genre_{it} + \beta_6 \cdot MovieMonth_{it} + \beta_7 \cdot PopularReview_{it} + \beta_8 \cdot CriticsReview_{it} + \beta_9 \cdot Sequel_{it} + \varepsilon_{it} \tag{4}$$

In this equation, the subscript "i" indicates individual films and within "t" years. $Release\ Year_t$ refers to year fixed effects for this equation between 2008-2019. β_0 refers the baseline logged US box office revenue without any changes in determinants of film revenue. β_1 is the effect changes in the ratio of female-to-male cast members has on the elasticity of US Revenue in terms of percent. β_2 refers to the elastic change between production budget and US revenue in percent. The Variable "Star Presence" is a dummy variable with four sub-categories, presence of a male star, presence of a female star, presence of both a male and a female star, and the presence of no star. β_3 is the percent change in US Revenue due to the presence of stars. β_5 refers to any elastic changes to US film revenue that are related to the Genre of a film, for which 5 dummy variables were created: Action, Comedy, Drama, Romantic Comedy and Thriller. Likewise, β_6 is the percent change on US box office revenue based upon the month a film was released. β_7 and β_8 respectively are the impact of Popular and Critics Review, both of which are the fractions against the maximum score. These coefficients are the impact of one-unit-increase or decrease of the popular and critics review of a film. β_9 is the elastic impact the film being a sequel. Finally, ε_{it} refers to the uncorrelated error term of the regression. The results of both the above equations are listed in table 6 and 7 below.

Table 6. Regression Results on log(Indian Revenue)

Variables	Coefficient	Std. Err	T-Value	P-Value
log(Production Budget)	1.2655***	0.2349	5.39	0.000
Gender Ratio	-0.0444	0.2001	-0.22	0.825
Star Presence (Baseline = No Star)				
Has Male Star	0.4773	0.3439	1.39	0.167
Has Female Star	0.7482*	0.4136	1.81	0.072
Has Male and Female Star	1.0856**	0.3820	2.84	0.005
Production Year (Baseline = 2009)				
2010	0.4517	0.3452	1.31	0.192
2011	0.1039	0.4562	0.23	0.820
2012	0.1834	0.3632	0.51	0.615
2013	-0.0474	0.6754	-0.70	0.483
2014	-2.6444***	0.5810	-4.55	0.000
2015	-2.0774***	0.4945	-4.20	0.000
2016	-1.7594***	0.5017	-3.51	0.001
2017	0.0427	0.3965	0.11	0.914
2018	-2.2740***	0.5436	-4.18	0.000
2019	-3.2843***	0.6668	-4.93	0.000
Genre (Baseline = Action)				
Comedy	0.7358*	0.3748	1.96	0.051
Drama	-0.0390	0.4162	-0.09	0.925
Romantic Comedy	0.3521	0.4029	0.87	0.383
Thriller	0.4679	0.6052	0.77	0.440
MovieMonth (Baseline = April)				
Jan	-0.4055	0.7703	-0.53	0.599
Feb	0.2322	0.8244	0.28	0.779
March	0.3509	0.8989	0.39	0.697
May	-0.5862	0.8655	-0.68	0.499
June	-0.2788	0.7303	-0.38	0.703
July	0.5006	0.7752	0.65	0.519
August	-0.6399	0.7542	-0.85	0.397
September	0.0968	0.7977	0.12	0.904
October	-0.0945	0.7808	-0.12	0.904
November	-0.2919	0.7435	-0.39	0.695
December	0.3250	0.7990	0.41	0.685
Popular Review	0.1500	0.1263	1.19	0.237
Critics Review	1.7833	1.4439	1.24	0.218
Sequel	0.0491	0.3710	0.13	0.895
Constant	-6.5356*	3.7710	-1.73	0.085
R^2	50.40%			
Observations	235			

*** p<0.01, ** p<0.05, * p<0.1

Table 7. Regression Results on log(US Revenue)

Variable	Coefficient	Std. Err	T-Value	P-Value
log(Production Budget)	0.9376***	0.2124	4.42	0.000
Gender Ratio	0.0661	0.1194	0.55	0.581
Star Presence (Baseline = No Star)				
Has Male Star	0.5305**	0.2380	2.23	0.027
Has Female Star	0.8429***	0.2148	3.92	0.000
Has Male and Female Star	1.1270***	0.2821	4.00	0.000
Production Year (Baseline = 2009)				
2010	0.4576	0.3588	1.28	0.204
2011	0.5874	0.3362	1.75	0.082
2012	0.5716	0.3501	1.63	0.104
2013	0.5926	0.4909	1.21	0.229
2014	0.2618	0.3579	0.73	0.465
2015	0.3335	0.3678	0.91	0.366
2016	0.6489*	0.3504	1.85	0.066
2017	0.4595	0.3893	1.18	0.239
2018	0.7349*	0.4035	1.82	0.070
2019	0.3691	0.4625	0.80	0.426
Genre (Baseline = Action)				
Comedy	0.1625	0.1991	0.82	0.415
Drama	-0.0134	0.1938	-0.07	0.945
Romantic Comedy	0.3611*	0.2128	1.70	0.091
Thriller	-0.0423	0.2490	-0.17	0.865
MovieMonth (Baseline = April)				
Jan	0.2101	0.3362	0.69	0.489
Feb	0.0675	0.3362	0.20	0.841
March	-0.2662	0.3920	-0.68	0.498
May	-0.3670	0.4517	-0.81	0.417
June	-0.0607	0.2329	-0.26	0.795
July	-0.0660	0.2629	-0.25	0.802
August	-0.2085	0.2457	-0.85	0.397
September	-0.1151	0.2644	-0.44	0.664
October	-0.0523	0.2464	-0.21	0.832
November	0.0271	0.2729	-0.03	0.976
December	0.1376	0.2656	0.52	0.605
Popular Review	0.2755***	0.065	4.24	0.000
Critics Review	2.334**	0.8483	2.75	0.006
Sequel	0.4164**	0.2097	1.99	0.048
Constant	-5.6023*	3.3621	-1.67	0.097
Adjusted R^2	64.28%			
Observations	235			

*** p<0.01, ** p<0.05, * p<0.1

All regressions used robust standard errors and underwent White's heteroskedasticity test. Figures 8 and 9 in the Appendix are the residual plots for both regressions and highlight limited heteroskedasticity. Further, the validity of the model was confirmed after running F-Tests for joint significance on all dummy variables and calculating variance inflation factors. All Variance inflation factors are shown in Table 8 of the Appendix. Due to all values being less than 5, I was able to assume no multicollinearity. While group testing dummy variables showed varied significance, individual categories proved to be significant. Thus, the categories remained in the regression.

Results and Findings

This section first explains results for Indian box office revenue regressions, then United States box office revenue, concluding with differences between both. Overall, these results indicate that there is no statistically significant relationship between the ratio of female to male cast members and box office revenue. However, closer inspection of the regressions highlight that viewing preferences of Bollywood films is incredibly heterogeneous.

Indian Box Office Revenue

The R-squared of this regression was 0.5040, indicating that 50.40% of all variances in the dataset was captured by the regression itself. There was not a statistically significant relationship between the overall ratio of female to male cast members and box office revenue. The resulting coefficient was very small at roughly -4%. However, the presence of a female star did have a significantly positive impact on box office revenue. This relationship is visualized in figure 6 below. On average, a film having only a female star with no male counterpart led to an increase in box office revenue of 75% and was significant at the 10% level. Having both a male and a female star lead to an increase in box office revenue of 108% and was significant at the 5% level. This indicates that the gender of movie stars plays a crucial role film box office success, and that presence of several superstars influences box office success. However, the presence of a male movie star does not dictate film success.

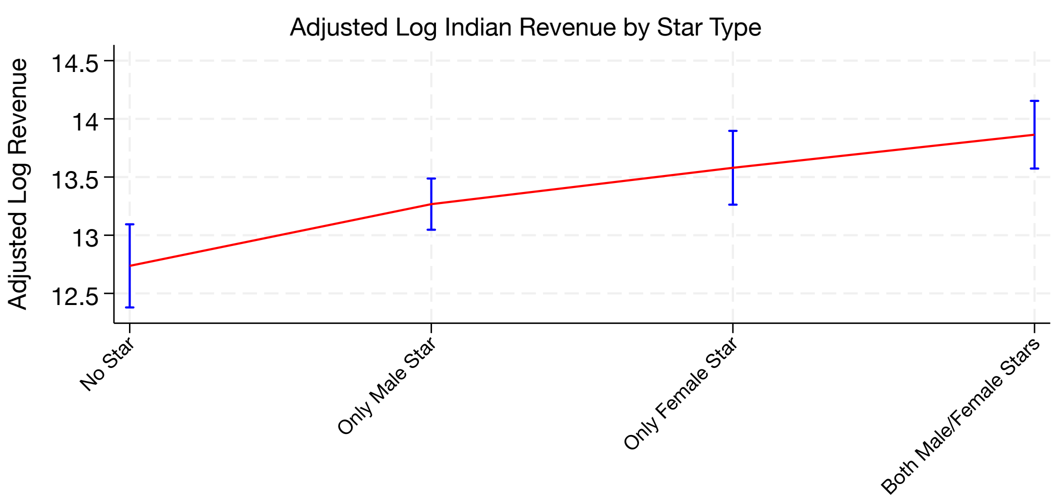


Figure 6. Margins Plot Based on Star Type, highlights increase in Box Office Revenue with female movie stars.

This differs from Aboli's study, who found male star presence to be more crucial. Notably though, her study does not consider the presence of several stars in a film (Aboli, 2022).

Other control variables were significant across the regression. In the Indian context, production budget emerged as the most influential box office predictor. With a coefficient of nearly 1.27, this suggests that on average, a 1% increase in production budget is associated with a 1.27% increase in box office revenue. This coefficient and corresponding p-value were the largest amongst all control variables. Ultimately, this aligns with prior studies in both the Hollywood (Prag & Casavant, 1994) and Bollywood contexts (Aboli, 2022). Certainly, a film's production budget has the largest economic impact on expected revenue.

Continuing with the quality indicators neither critic's reviews nor popular review significantly predict box office revenue, sharing similar insignificance. It should be noted that the critics response to films did increase predicted box office revenue by almost 178%. This finding ultimately highlights the disconnect between Dastidar and Aboli's studies. Both find different impacts between critics and popular review, indicating their relatively weakness as predictors of box office success.

By group testing the three dummy variables in this study – genre, production year, month of release, and star presence – we found that all these variables are crucial in predicting a film's overall success. However, only a few categories were individually significant.

Of the genre variables studied - Action, Comedy, Drama, Romantic Comedy and Thriller/Suspense - only Comedy was significant in the Bollywood context at the 10% level. This contradicts findings by Aboli, who found Drama and Action genres to best predict box office revenue. It is important to note that my dataset comprised equal distribution across Action and Comedy films, unlike Aboli which skewed towards Comedy films. The overall relationship was estimated to be 74% higher revenue when considered a "Comedy" in the Indian context.

None of the individual Month dummy variables were statistically significant; however, the months of February, July, September, March and December showed a weakly positive relationship with box office revenue. Although not significant at any p-level, these months are unique as they align with traditional Indian holidays and corresponding film release schedules (Aboli, 2022). Of the year fixed effects, 2014, 2015, 2016, 2018 and 2019 were significantly negative on box office revenue. This could be due to streaming; however, further research is needed.

United States Box Office Revenue

The adjusted R-square of this regression is 0.6428 indicating that 64.3% of all variances in the data is captured by the regression. The ratio between female to male cast members did not have a significant relationship on US box office revenue. Despite this insignificance, the approximate sign of the relationship is positive, with its coefficient at a roughly 7% revenue increase for every one-unit increase in female to male ratio. The gender of the star presence within a film was significant. Every star presence indicator had a significantly positive relationship on box office success. Figure 7 visualizes this relationship.

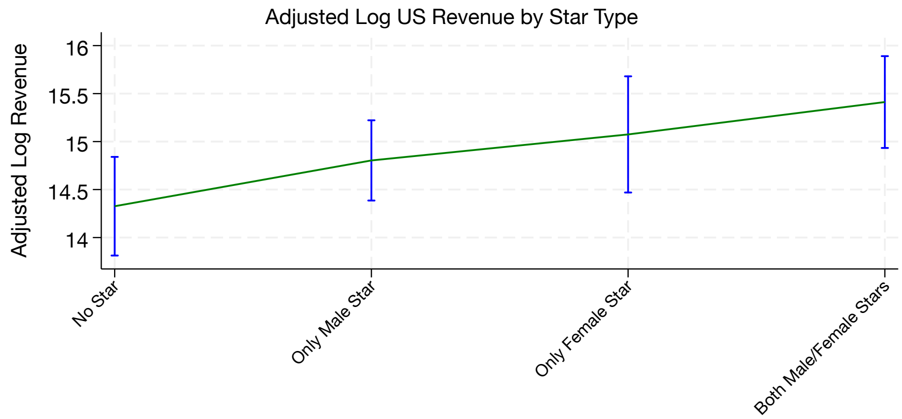


Figure 7. Margins Plot Based on Star Presence, highlights increase in Box Office Revenue with female movie stars.

The presence of both male and female stars generated the largest positive regression coefficient, indicating nearly a 113% increase on box office revenue. This finding was significant at the $p < 0.01$ level. Furthermore, films with solo female stars are associated with larger box office revenue by nearly 84%. This increase approximately 30% higher than the presence of a solo male star.

Production Budgets also predicted box office revenue, with a 1% increase in production budgets signifying a 0.94% increase in box office revenue. This is common within film economics (McKenzie, 2022). The presence of sequels also had positive and significant relationships with expected box office revenue. We found that films which were sequels had an estimated 42% positive increase in box office revenue. This indicates that within the United States market, sequels are strong indicators of box office success.

Despite completing an F-Test for joint significance and finding that all group dummy variables were significant, only a few individual categories showed statistical significance. Of the genre variables, Romantic Comedies positively influence box office revenue by 36%. Of the month and year fixed effects, only 2016 and 2018 were significant. Due to joint significance both were kept in the final regression equations and results.

Differences In Indian and American Markets

The direction and significance of gender-based cast composition was different in Indian and American markets. Both findings were statistically insignificant, however the United States relationship saw stronger relationships (p=0.581) versus the Indian market (p=0.825). This notable difference, indicating that the predictive power could be more influential in the United States context. Furthermore, the directionality of the findings positive in the United States and negative for India. This distinction is important, as it could illuminate distinct viewing preferences across the Indian diaspora. Within an American context, Indian American individuals are typically higher earners and younger (The Global List, 2025) compared to Indian nationals. This could influence viewing patterns as these populations could prefer films reflecting lived experiences. Romantic Comedies had the strongest significance in the American context which might have more gender parity, especially as the storylines are often about both men and women finding partnership.

The star presence variable had different impact on box office revenue between markets; in the United States all star variables were significant. In India only female stars and both male/female stars showed impact on box office revenue. The magnitude of this impact was similar in both markets. However, female stars significantly impact box office revenue compared male stars. This may point to female movie stars being novel to audiences. Further, exposure to female stars may be influenced by external brand partnerships and advertisements. It is possible these non-film related enterprises increase connection between audiences and female movie stars. In the United States, audiences might consider Indian superstars as equally important, especially as various brand engagements are not shown on traditional forms of American media. Additionally, in the United States there are more film options from Hollywood. This result may be due to substitutability between Bollywood and Hollywood films for American audiences. Consequently, Bollywood films may require more star power to captivate audience attention in crowded markets.

Furthermore, the overall lack of significance of gender composition is likely attributed to other variables having stronger impact on box office revenue, such as production budget. A larger production budget correlates with a higher box office revenue, and a greater ratio of female to male cast members is correlated with lower production budget. While not the focus of this study, an area for future research is the differences between budgets given to women in Bollywood.

There were numerous external differences in results between the United States and Indian box office contexts. This highlights composite differences in film consumption and cultural practices. To begin, neither “critics’ review” nor “popular review” variable was a significant factor contributing to Indian box office revenue, whereas both were significant in the United States. This aligns with the findings of Aboli and Basuroy, respectively. Aboli finds popular reviews to not impact Indian Box Office Revenue. (Aboli, 2022) On the other hand, Basuroy found that critics’ reviews and popular reviews signify positive box office revenue in the United States Box Office (Basuroy et al, 2003). This indicates differences in the preferences of film audiences in the United States and India, with American film viewers considering a film’s quality in addition to the presence of a movie star.

Another key distinction lies in the magnitude of impact production budget on a film’s expected revenue. Compared to the United States context, the Indian context has significantly higher expected box office revenue based on production budget. This could be explained by a few reasons: First, production budget information may not inform demand as strongly for American viewers unfamiliar with specific budgetary information. In other words, a “Bollywood” film is considered the same regardless of size. Second, production budget estimates in India could influence marketing and advertising costs domestically. In the United States, the same marketing spend would not influence consumption as those channels would be cannibalized by Hollywood features. That said, production budget is clearly an influential factor in expected revenue of a film, regardless of country.

Unlike in the Indian context, sequels were statistically significant in the American context. This aligns with the findings of Walls (Walls, 2009) and others on the impact of sequels. Walls finds that once an audience is primed for, and enjoys a film, its sequel does well at the box office. This diverges from the Indian context, building on work done by Dastidar and Elliott, who found that sequels did not impact Indian film revenue in India (Dastidar, et al., 2020).

Limitations

This study expands film economics in two keyways. First, we utilize new, verifiable data in our econometric analysis of Bollywood film review. This verification is crucial as it legitimizes prior work which typically relies on online sources. Second, we continue to look at the overall impact gender and representation play into the filmmaking process. Due to this novelty, there are a few limitations.

First, the dataset used in this study is small with a sample size of 235 and is subject to the larger film data issues within Bollywood. The overall lack of centralization in the filmmaking apparatus in Bollywood often means that limited standardization practices for reporting box office and production data. In practice, many 3rd party data companies do not have access to wide swaths of information. As such, my dataset does not include all releases within the 10-year period. Furthermore, the results of this study may be impacted by data accuracy and measurement errors. As explained in prior sections, data relevant to the study of Bollywood films has been minimal. OpusData for example, relies on box office receipts and data scraping. While the legitimacy of this data is verifiable, there is no method to validate the estimations of production budgets or the impact of non-legitimate actors throughout the production process. Future research and policy must ensure that data within Bollywood is accessible and verifiable. With that accomplished, this study's methodology can be replicated to continue to explore the relationship between gender representation and film outcomes. Additionally, as is standard across film economics, skewness of data has impacted overall results. This has been rectified in part due to the log-transformation of key data; however, this does not account for all skewness.

Second, production budget estimations do not include marketing and advertising expenditure in the lead up to a film. Production budget has been used in prior literature to serve as a stand-in for these costs, with the assumption being that higher production budgets indicate more spending on marketing. However, exact marketing and advertising expenditure figures are proprietary. As a result, there is no meaningful way to allocate portions of production budget spent on marketing and advertising. Along similar lines, a given star may have their own personal marketing budget through individual social media accounts, which influences the success of a film but is not financed by the studio. I have accounted for this by controlling for star presence. Due to these complexities, it is impossible to definitively conclude that a higher production budget correlates with larger marketing spend.

Third, this study highlights the female-to-male ratio as a proximate gauge for female representation. However, there may be other gender norms and influences which dictate the relationship between ratio and film revenue. While I have employed year fixed effects to mitigate influence of gender or other economic factors which may change over time, there may be other gender norms at play. To that end, the overall impact of things such as social media and the "Me Too" movement are not analyzed in this study. After these movements there might have been more attention paid by audiences on representation which may have been flattened by controlling for all years.

Fourth, with the rise of social media and streaming, it is possible that box office revenue does not fully capture the impact of gender representation on new media. While box office revenue was used in this study to adhere to prior literature, streaming and online viewing has changed viewing preferences across the globe. Further research must be done in terms of mitigating the overall industry loss to better understand how streaming inhibits or adds to the market.

Conclusions

This study focused on the impact of gender, particularly the ratio of female to male cast members, on Bollywood box office revenue, highlighting differences between United States and Indian box office revenue. While statistically insignificant, there is a difference in the relationship between the ratio of a film's cast and its box office revenue. Furthermore, movie stars had an unquestionable impact on box office revenue both in India and United States. This highlights audiences' strong reaction to actresses they enjoy. It is possible that with more financial resource's films will continue to see strongholds at the box office. Additionally, differences in impact highlight larger distinctions between American and Indian consumers.

This study indicates that as more Bollywood films are presented on a global stage through streaming services, the genre and gender representation could increase global film viewership. Nonetheless, this relationship must be further studied to understand these nuances with evolving with different gender norms, expressions, and film watching preferences. Additionally, streaming might directly influence box office revenue mid to late 2010s. Future research should consider the most significant year streaming impacted Bollywood as this might increase demand for diverse gender representation.

Though this study had data limitations, as do many focusing on Bollywood, the overall conclusions regarding determinants of box office revenue add to overall literature. Future policy must ensure that box office and other film data is shared throughout the film apparatus, as the limited access impacts research goals.

Importantly, my study highlights differences in gender representation by looking at both narrative representation and female star presence. When measured in terms of ratio of male to female cast there was not a significant impact. However, when measured with respect to stars there was a significant gender-based impact. There are numerous reasons for this including the diluting cast impact on demand. In other words, all members of on-screen talent are not considered the equal by audience members. Future research must continue analyzing gender through distinct lenses such as composition versus narrative analysis. Especially when influencing production budget and sizing, representation shapes the content we consume. Finally, it is crucial to demand stories and movie stars which reflect lived experiences such that we are all able to watch ourselves on screen.

<u>**Reference List**</u>

- Aboli, Z (2022). *Box Office Success of Bollywood Films. An analysis of influencing factors*. Creative Industries in India, **10**, 203-226.
- Basuroy, Suman & Chatterjee, Subimal & Ravid, S. Abraham. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. Journal of Marketing - J MARKETING. 67. 103-117. 10.1509/jmkg.67.4.103.18692.
- Hao, Bingyu. (2023). The Analysis of the Factors that Influence the Film Revenue. Highlights in Science, Engineering and Technology. 47. 154-159. 10.54097/hset.v47i.8184.
- Bechdel, A. (1986). *Dykes to watch out for*. Firebrand Books.
- Berg, J., & Raddick, M. J. (2016). First You Get the Money, Then You Get the Reviews, Then You Get the Internet Comments: A Quantitative Examination of the Relationship Between Critics, Viewers, and Box Office Success. Quarterly Review of Film and Video, 34(2), 101–129. https://doi.org/10.1080/10509208.2016.1156426
- Bollywood Hungama (2019). *Top 100 Hindi Bollywood Celebrities* [Data set]. https://www.bollywoodhungama.com/celebrities/top-100/
- Brewer, S., Kelley, J., & Jozefowicz, J. (2009). A blueprint for success in the US film industry. Applied Economics, 41, 589-606
- Chisholm, D.C. Two-Part Share Contracts, Risk, and the Life Cycle of Stars: Some Empirical Results from Motion Picture Contracts. Journal of Cultural Economics 28, 37–56 (2004). https://doi.org/10.1023/B:JCEC.0000009808.60007.ea
- Collins, A., Hand, C., Cnell, M. (2002). *What Makes a Blockbuster? Economic Analysis of Film Success in the United Kingdom.* Journal of Managerial and Decision Economics, **23(**6), 343-354.
- Dastidar, S.G., Elliott, C. The Indian film industry in a changing international market. J Cult Econ 44, 97–116 (2020). https://doi.org/10.1007/s10824-019-09351-6
- Davids, E. (2023). The Evolution of Film Genres: A Comparative Analysis of Hollywood and Bollywood. rt and ociety, 2(4), 15–34. retrieved from https://www.paradigmpress.org/as/article/view/736
- De Vany, A. and Walls, D. (1999) Uncertainty in the movie industry: does star power reduce the terror of the box office? Journal of Cultural Economics 23: 285–318. Also published in De Vany, A. (ed.) (2004) Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry. London: Routledge.
- De Vany, A. (2003). Hollywood economics: How extreme uncertainty shapes the film industry. Routledge.
- Delaney-Bhattacharya, A. (2020). *The 2010s: Reflecting on a decade of change in Bollywood*. – The Diplomat. https://thediplomat.com/2020/01/the-2010s-reflecting-on-a-decade-of-change-in-bollywood/

- Einav, L. (2007) Seasonality in the US motion picture industry. *The RAND Journal of Economics*, **38**(7), 127-145.
- Elberse, A., & Eliashberg, J. (2003). Demand and supply dynamics for sequentially released products in international markets: The case for motion pictures. *Marketing Science* 2003, vol. 22, issue 3, 329-354.
- Eliashberg, J., & Shugan, S. M. (1997). Film Critics: Influencers or Predictors? Journal of Marketing, 61(2), 68–78. https://doi.org/10.2307/1251831
- EY. (March 12, 2024). Value of the media and entertainment industry in India from financial year 2019 to 2023, with estimates until 2026 (in billion Indian rupees) [Graph]. In Statista. Retrieved March 25, 2025, from https://www.statista.com/statistics/235784/india-value-of-the-media-and-entertainment-industry/
- Feng, F., & Sharma, R. (2016). Modeling the main determinants of movie sales: An econometric study of Chinese marketplace. Journal of Reviews on Global Economics, 5, 190–209.
- Feng, G. C. (2017). The dynamics of the Chinese film industry: Factors affecting Chinese audiences' intentions to see movies. *Asia Pacific Business Review*, **23**, 658–676

- Ginsburgh, V., Weyers, S. On the Perceived Quality of Movies. Journal of Cultural Economics 23, 269–283 (1999). https://doi.org/10.1023/A:1007596132711

- Hermosilla, M., Gutierrez-Navratil, F., & Prieto-Rodriguez, J. (2018). Can emerging markets tilt global product design? Impacts of Chinese colorism on Hollywood castings. *Marketing Science*, 37(3), 356–381.

- Heywood, J.S., Izquierdo Sanchez, S. & Paniagua, M.N. The Hollywood gender gap: the role of action films. *J Cult Econ* (2024). https://doi.org/10.1007/s10824-024-09514-0

- International Movie Database (2009-2019). *Individual Film Information* [Data Set]. www.imdb.com
- Hoffman, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2017) Empirical generalizations on the impact of stars on the economic success of movies. *International Jorunal of Research in Marketing* Volume 34, Issue 2, Pages 442-461.
- *Indywood: Landscape of the indian film industry: Article*. Deloitte India. (2016, January 4). https://www2.deloitte.com/in/en/pages/technology-media-and-telecommunications/articles/indywood-indian-film-industry.html
- Jha, L. (2023, January 5). Bollywood, Box Office, and Beyond: The Evolving Business of Indian Cinema.

- Khadilkar, K., KhudaBukhsh, A., & Mitchell, T. (2022) Gender bias, social bias, and representation in Bollywood and Hollywood. *Patterns*, Volume 3, Issue 4, 8 April 2022, Pages 100486. https://doi.org/10.1016/j.patter.2021.100409
- Kim, D. H., & Kim, S. (2018). A comparative study: Performance of Hollywood and Korean sequel films in Korea. Journal of Media Economics, 31(1–2), 36–49.
- Lim, V. (2016, August 10). Increased foreign revenue shares in the united states film industry: 2000-2014. Duke Journal of Economics. **28**
- Lindner, A. M., Lindquist, M., & Arnold, J. (2015). Million dollar maybe? The effect of female presence in movies on box office returns. Sociological Inquiry, 85, 407–428.
- McMahon, J. (2023). Star Power and Risk: A Political Economic Study of Casting Trends in Hollywood. Quarterly Review of Film and Video, 41(8), 1414–1447. https://doi.org/10.1080/10509208.2023.2215355
- Maindargi, S. (2020, May). *An Empirical Study of Entertainment Industry and its impact on Indian Economy with Spicial Referace to Solapur City (Maharastra).* International Research Journal of Modernization in Engineering, **02**(05).
- Mazières, A., Menezes, T. & Roth, C. Computational appraisal of gender representativeness in popular movies. Humanit Soc Sci Commun 8, 137 (2021). https://doi.org/10.1057/s41599-021-00815-9
- Michelle C. Pautz (2015). "'Argo' and 'Zero Dark Thirty': Film, Government, and Audiences." *PS: Political Science & Politics,* 48, pp 120-128. doi:10.1017/S1049096514001656.
- McKenzie, J. (2012), THE ECONOMICS OF MOVIES: A LITERATURE SURVEY. Journal of Economic Surveys, 26: 42-70. https://doi.org/10.1111/j.1467-6419.2010.00626.x
- McKenzie, J. (2022). *The economics of movies (revisited): A survey of recent literature.* Journal of Economic Surveys. https://doi.org/10.1111/joes.12498.
- Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. Review of Economic Studies, 78, 356–393.
- Nguyen, J. (2023, Jan 16). *The economic case for equality in screenplays: The Bechdel test, female dialogue and box office revenue*. Journal of Screenwriting, **14**(2), 173-190.
- Niraj, R., & Singh, J. (2015). Impact of User-Generated and Professional Critics Reviews on Bollywood Movie Success. Australasian Marketing Journal, 23(3), 179-187. https://doi.org/10.1016/j.ausmj.2015.02.001 (Original work published 2015)
- Palia, D., Ravid, S. A., & Reisel, N. (2008). Choosing to cofinance: Analysis of project-specific alliances in the movie industry. The Review of Financial Studies, 21(2), 483–511.

- Pandey, G. (2022, September 12). Rising crimes against Indian women in five charts. The British Broadcasting Company. https://www.bbc.com/news/world-asia-india-62830634
- Patil, V., Date, H., Kumar, S., Lim, W.M. and Donthu, N. (2022), "The making of box-office collection: qualitative insights from Bollywood", Marketing Intelligence & Planning, Vol. 40 No. 8, pp. 1010-1023. https://doi.org/10.1108/MIP-07-2021-0238

- Rathore, M. (2025, March 06). Women in India. Statistia https://www.statista.com/topics/5220/women-in-india/#topicOverview

- Reinstein, D.A.; Snyder, C.M. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. J. Ind. Econ. 2005, 53, 27–51

- Smith, Stacy & Pieper, Katherine & Choueiti, Marc. (2010). Assessing Gender-Related Portrayals in Top-Grossing GRated Films. Sex Roles. 62. 774-786. 10.1007/s11199-009-9736-z.

- Treme, J. (2010). Effects of Celebrity Media Exposure on Box-Office Performance. Journal of Media Economics, 23(1), 5–16. https://doi.org/10.1080/08997761003590457

- Treme, J., Craig, L. A., & Copland, A. (2018). Gender and box office performance. Applied Economics Letters, 26(9), 781–785. https://doi.org/10.1080/13504851.2018.1495818

- The Times of India. (2009-2019) [Data Set]. Film Critic's Review. *The Times of India.* http://timesofindia.indiatimes.com/

- The Lab Mag. (2018, March 12). Bollywood analysed: Films that qualify the Bechdel test. The Ideas Lab. https://theideaslab.com/bollywood-analysed-films-qualify-bechdel-test/

- The Global List (2025, January 14). Indian americans: A survey data snapshot. *The Globalist*. https://www.theglobalist.com/united-states-india-immigration-workers-salary-economy/
- Prag, J., & Casavant, J. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics*, **18**(3), 217–235.
- Shackleton, L. (2023, May 16). *India recorded second-biggest box office year ever in 2022 with $1.28B, but admissions still down on pre-pandemic levels*. Deadline.

https://deadline.com/2023/02/india-box-office-second-biggest-year-kgf-chapter-2-rrr-avatar-1235261274/

- Valentowitsch, J. (2022). Does female screen presence pay off at the box office? An exploratory analysis with special emphasis on the Bechdel test. *Journal of International Business and Economics., 22*(1), 40–69.
- Valentowitsch, J. Hollywood caught in two worlds? The impact of the Bechdel test on the international box office performance of cinematic films. Mark Lett 34, 293–308 (2023). https://doi.org/10.1007/s11002-022-09652-5
- Walls, W. D. (2009b). Screen wars, star wars, and sequels. *Empirical Economics*, **37**, 447–461.
- Wall Street Journal (2018). *INR/USD* [Data set]. https://www.wsj.com/market-data/quotes/fx/INRUSD/historical-prices
- Wühr P, Lange BP and Schwarz S (2017) Tears or Fears? Comparing Gender Stereotypes about Movie Preferences to Actual Preferences. Front. Psychol. 8:428. doi: 10.3389/fpsyg.2017.00428
- Palia, D., Ravid, S. A., & Reisel, N. (2008). Choosing to cofinance: Analysis of project-specific alliances in the movie industry. The Review of Financial Studies, 21(2), 483-511.
- Yang, L., Xu, Z., & Luo, J. (2020). Measuring female representation and impact of films over time. *ACM Transactions on Data Science,* 1(4), 1-14.
- Zaidi, Saher. (2021) Changing portrayal of a woman in Bollywood cinema: special reference of movies of Vidya Balan**.** Retrieved from https://doi.org/doi:10.7282/t3-xh4s-vz43

Appendix

Residual Plots

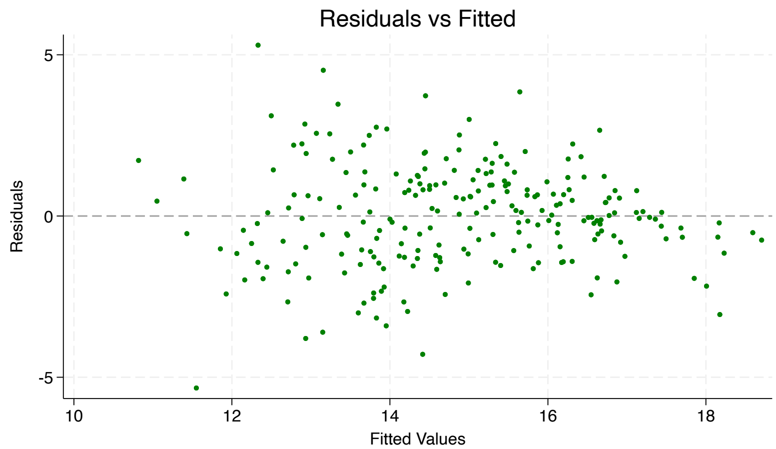


Figure 8. Residual Plot log(Indian Revenue), indicates slight heteroskedastic error however, this is common.

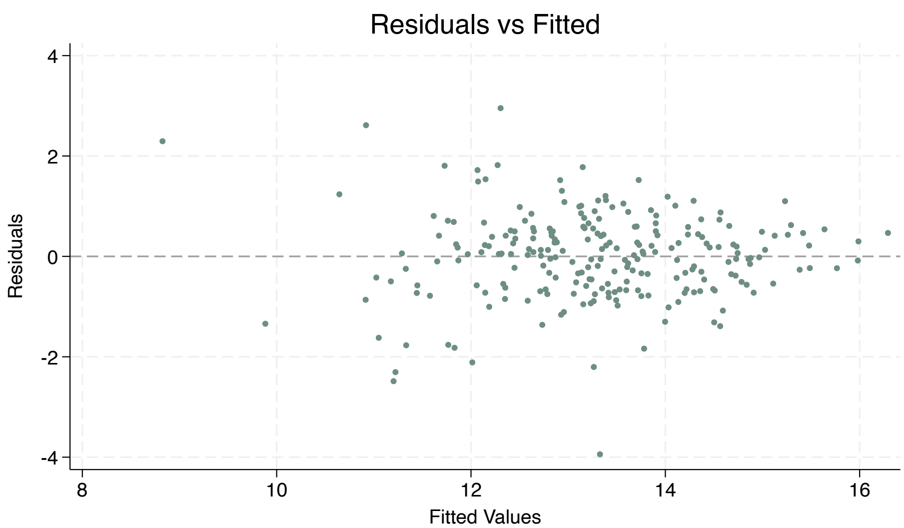


Figure 9. Residual Plot log(US Revenue) indicates no visual presence of heteroskedastic error.

Multicollinearity

Variable	VIF	1/VIF
log(Production Budget)	2.72	0.367755
Gender Ratio	1.2	0.830969
Male Star	2.04	0.489711
Female Star	1.47	0.680019
Both Star	2.57	0.389817
2010	1.80	0.55687
2011	1.81	0.553007
2012	1.85	0.539913
2013	1.71	0.585784
2014	1.85	0.539846
2015	2.36	0.424289
2016	2.04	0.490987
2017	1.91	0.525225
2018	1.86	0.536725
2019	1.91	0.524185
Comedy	1.96	0.509713
Drama	2.24	0.445728
Romantic Comedy	1.98	0.505593
Thriller/Suspense	1.64	0.608599
Jan	2.62	0.381919
Feb	2.85	0.350796
Mar	2.59	0.385721
May	2.10	0.476233
Jun	2.72	0.367739
July	2.83	0.353656
Aug	2.92	0.341918
Sept	3.29	0.304396
Oct	2.88	0.347398
Nov	2.60	0.384454
Dec	2.58	0.386966
Popular Review	2.22	0.450143
Critics Review	2.06	0.485413
Sequel	1.28	0.780763

Table 8. Variance Inflation Factors all within accepted range, no presence of Multicollinearity.